Intrusion Detection by Intelligent analysis of data across multiple gateways in real-time.

Joel Scanlan, Samuel Lorimer, Jacky Hartnett and Kevin Manderson

School of Computing, University of Tasmania
Sandy Bay Tasmania

jdscanla@utas.edu.au, sam.lorimer@bigpond.com, J.Hartnett@utas.edu.au, kevin.m@webos.com.au

Keywords: Intrusion Detection, Firewall, Multiple Gateway, Analysis

Abstract – Current firewalls and intrusion detection systems are generally designed to protect a single gateway in order to provide protection for machines residing behind the gateway on an internal network. When considering a network incorporating multiple gateways across a range of IP addresses exposed to the Internet, interesting data can be gathered with regard to the types of scans occurring across these gateways from the outside. The validity of using a central server to amalgamate, reduce and analyse the log files of each gateway is investigated in order to examine the activities of the scans across multiple gateways and port numbers. The results from this analysis can then be used to act against an attack through heuristic driven rule creation.

INTRODUCTION

In the last few decades computers, and their attached networks, have spread throughout the globe to the stage where they are now pervasive in our everyday lives. However while these technological leaps have arguably made our lives easier and our jobs more efficient they have not necessarily made our data (of whatever type or content) more secure. Work, home and government networks are now all connected through the internet, leaving them open to attack, whether the attack's origin is within the same city as the server storing the data, or is half a world away on a different continent.

A Trend Micro [1] survey of 500 corporations, government agencies, financial institutions, medical institutions, and universities revealed that 85% of them had suffered a security breach in the preceding 12 months. Everyone in the networked world is at risk of those who wish to cause harm to computer systems.

In an effort to counter these threats network infrastructure in the form of Firewalls and Intrusion Detection systems are implemented. Firewalls act as a perimeter defence around a network, blocking traffic that is not allowed to enter (or exit) a network. Intrusion detection systems are used to alert administrators to threats that have made it through the firewall onto the network, often producing an automated response or action.

Current intrusion detection systems and firewalls are designed to operate on a single gateway as an individual location separate to the other components of the network (or as an individual listening sensor monitoring network traffic). This results in them not having access to the contextual information of what is occurring to the network as a whole; which frequently involves having other gateways to other networks and the internet.

The following paper will describe the analysis processes and results of an examination of an Audit log that was amalgamated from multiple gateways across an entire class C IP address range. Attacks that are being carried out against the network as a whole will be responded to, pre-empting the attack's continuation against other gateways on the network.

AUDIT LOG ANALYSIS

The most fundamental element in almost any intrusion detection system is the presence of a log

file. Audit log files record the events that occur on a computer system, along with a time-stamp and other identifiers such as the user or IP address. Without this critical information it is impossible to know what operations have been performed on the system. The log files used within this study were from the gateways (entry points) to a network.

Audit log files have primarily been used in the past to analyse how an attack occurred upon a system after it has finished [2]. Clifford Stoll [3] in his book *The Cuckoo's Egg* details how he tracked a series of attacks by a hacker on and through his system by printing out his activities and laboriously manually analysing what the hacker had done. During the 1980's audit logs changed from being massive mounds of weeklong printouts to being stored electronically on the system. Developments in pattern matching techniques allowed for automated analysis of electronically stored audit logs [4]; these were the first intrusion detection systems.

During the 1990's advances in computing power enabled the analysis of audit logs to occur in real-time, thus allowing these intrusion detection systems to respond immediately to attacks [5]. The events which usually result in a log being created in most systems entail identification and authentication mechanisms, creation, deletion or modification of files and directories, network activity and administrative activity relating to processes and account creation [6].

Before analysis can occur the data stored within the file is filtered, discarding information that is irrelevant to the analysis. Feature extraction is a further process of log reduction; it examines the log file entries, extracting specific relevant information and again discarding the remainder. These methods facilitate fast efficient extraction of audit log data.

There are two main methods by which attacks are discovered upon a system through Log File Analysis: Anomaly Detection and Signature Detection.

Anomaly detection ID systems require a profile of each user or user group to be made to enable the system to "learn" what comprises normal behaviour [7, 8]. The behaviour model is then compared to user actions upon the system, searching for behaviour that does not fit the model; this behaviour is then classed as abnormal behaviour and treated as an intrusion.

Anomaly detection is broader then just mapping profiles of human usage. It is also applicable to processes and network access or usage [9]. Network traffic analysis also yields profiles of normal usage that can be used in monitoring network traffic for anomalies and thus to detect attacks.

Signature detection searches audit logs for known attacks, matching malicious behaviour to pre-defined signatures. Signature or misuse detection has a database of attack signatures against which it can compare network event patterns in order to discover an attack. This results in signature detection systems being able to be operational directly after they are installed without the need for any training of the system [8].

Signature based intrusion detection is significantly more computationally efficient then anomaly based detection per item of knowledge as it does not need to create matrices for each system activity [10]. However, Brox [11] comments that signature detection has a flaw in that it requires a signature for a given attack to be able to be detected, and in some instances this is a case of waiting for an attack to occur, to then be able to make a signature to protect against it. Existing signature based intrusion detection systems examine audit logs within the context of a single gateway, and do not therefore protect systems from a signature that in actuality is spread across several gateways. It is this shortfall that is examined in this paper.

MULTIPLE GATEWAY ANALYSIS

Current Intrusion Detection Systems using signature or anomaly detection (or indeed combinations of them both) work effectively upon a single gateway or network device, however they lack the context of what is occurring across the entire network. Network infrastructure need to be contextually aware to be truly effective and efficient. An example of this can be seen in early packet filtering firewalls that lacked the contextual information of session data, and thus somewhat needlessly and laboriously filtered each packet within a session, ignoring the previous conclusion that the packets were not malicious in content. Likewise, multiple gateways across a single network could each be being (trivially) attacked simultaneously, each largely ignoring the attack, when in reality the attack is occurring across the whole network and is of a serious concern.

To be able to efficiently monitor and act upon such attacks a centralized analysis module needs to operate, having access to the complete audit logs of each gateway or sensor upon the network so as to preserve the network context. The Log amalgamation allows for signature and anomaly detection methods to be implemented network wide, thus allowing for a unified defence across the multiple gateways of the network.

IMPLEMENTATION

The implementation utilises actual audit data from a gateway range that consists of multiple remote gateways along with the central server (ns1). Ns1 is bound to the IP addresses of almost a complete C-class running from 0 to 252 in the last octet. This acts as an excellent range of 'virtual' consecutive gateways, as it will appear externally as though there are 253 separate machines when really each IP address will report to the same machine. The audit log still reports which IP address within this range was probed, meaning that it is possible to analyse the data from this single machine as though it were 253 separate gateway machines (Figure 1).

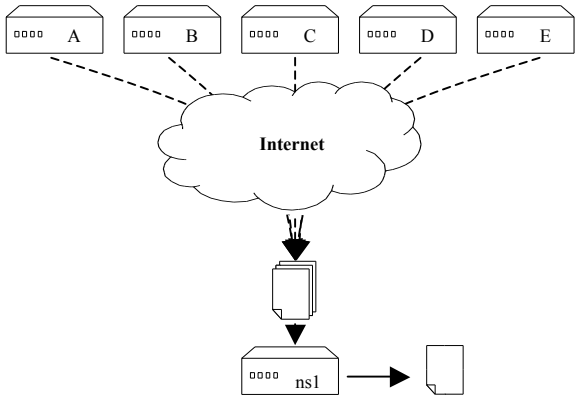

Fig. 1. Audit Log Amalgamation.

The majority of the results discussed within this paper were gathered from port probes that covered this IP range. One of the goals of the system is to develop an effective threshold level of probes which if passed would result in action being taken to protect the network as a whole.

The system developed to obtain results consists of two separate modules – an analysis module and a tracking module. A database is utilised to store the processed data.

The analysis module is developed to maintain an overall image of the state of the system at any given time. It analyses every incoming log entry, or an archived log file, and updates a table in the database which specifies which source IP addresses may be of interest. Rather than keeping a database entry for every instance of each IP within the log, it only makes one entry for each individual IP address and then increments a count for all subsequent instances along with additional Boolean information indicating if an IP has scanned more then a single gateway, or port.

This count that is recorded, is the value used by a that the threshold or heuristic that dictates how long an IP address can probe the network before it is banned upon all the gateways; this enables the analysis and tracking module to have a window of opportunity to gather the required information to

assess whether or not an IP is conducting a multiple gateway attack and set a Boolean multiple gateway attack value.

The tracking module is developed to follow the exact activities of individual IP addresses which have been deemed to be performing suspicious, or potentially interesting, types of scans across multiple gateways or port numbers. It creates a database entry for each piece of activity within the log file from the IP addresses that it is tracking. This results in a highly verbose, yet extremely comprehensive, record of the activities of these particular IP addresses. This data can then be analysed to gain an insight into the methods used for these scans.

The aim of these two modules is to track the activities of IP addresses across the gateway range in an attempt to recognise any patterns or methods of attack that are currently being overlooked by the existing security infrastructure examining them each singularly.

RESULTS

The results that have been ascertained fall into two categories, one for each module, Analysis Results and Tracking Results.

Analysis Module Results

The analysis module records a set of information on each IP which probes one of the gateways on the network; it records the number of times this IP has probed the network, whether or not it has probed more then one gateway, what port it probed on and whether or not it has probed multiple ports. When examining these data sets the following simple statistic table can be identified (Table 1).

During the 10-day study period (September 1st till 10th 2003) 6766 individual IP addresses probed the gateways on the network; of these 776 (11.5% of the total) probed more then one of the gateways. This demonstrates that there is a sizeable risk to systems from malicious users who are approaching gateway access at a network wide level, thus justifying central processing of audit data to retain network context. Realising that the presence and mode of attack is present, however, is only the first step to combating the problem itself. Detecting when these attacks are taking place with reasonable efficiency is the true goal of the Analysis module.

	Single Gateway	**Multiple Gateways**
Source IP Addresses	5990	776
% of Total	88.5	11.5

Table 1. Individual IP address Statistics.

Upon further examination of the data collected by the analysis module, it is possible to group the source IP addresses based on the total number of probes sent. Fig. 2 (over page) shows that the vast majority (83%) of source IP addresses sent only 3 or fewer probes against the network, indicating that perhaps it could be an appropriate level to test a threshold level heuristic. There are also slight increases at 6 and 9 also which were tested as threshold levels. This heuristic was then used in conjunction with the Boolean value stored for whether or not an IP has probed more then one gateway to classify source IP's, creating a less coarse heuristic rule set.

The results showed that at a threshold level of 3 only 3.2% of IP's were classified as potentially performing scans on multiple gateways. With the optimum of 11.5% to get all potential malicious probes it is a relatively poor result. By comparison, as the threshold level was increased to the levels of 6 and 9, the result returned were 8.3% and 10% respectively. These results were much more acceptable, however not quite at the levels desired to achieve an acceptable efficiency.

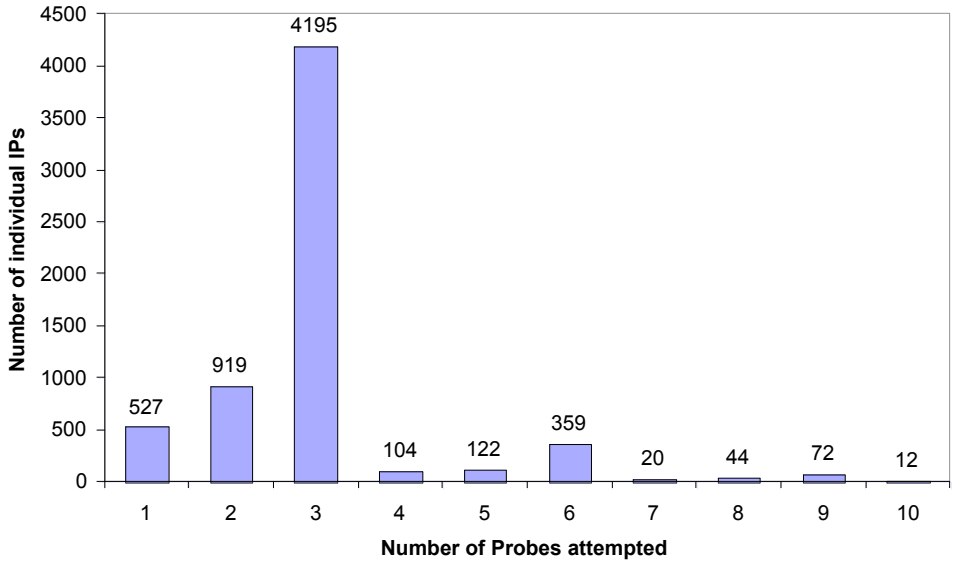

Fig. 2. Distribution of the number of port probe attempts

Fig. 3 illustrates the final distribution after testing several other threshold levels. The optimum level of efficiency was found to be when operating an 11 probe threshold.

Tracking Module Results

The tracking module produced very interesting results by using the flags that were triggered by the analysis module as a guide to which source IP addresses were worth tracking. The graph in Figure 4 depicts the ten days of Tracker activity on all source IP addresses that probed the gateway array. More than 40,000 probes were received in total in order to statistically analyse the scan activity on the gateway array as shown in Fig. 4. The scans across multiple gateways from one or more source IP addresses appear as vertical bars within this graph because of the large scale of the x-axis.

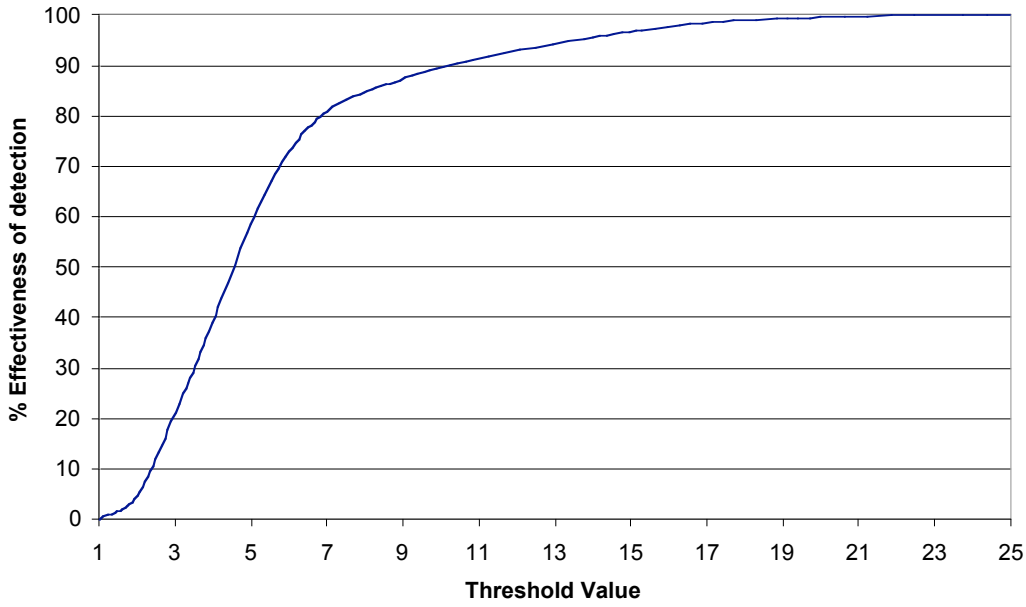

Fig. 3. Effectiveness of Detecting IP address probing Multiple Gateways

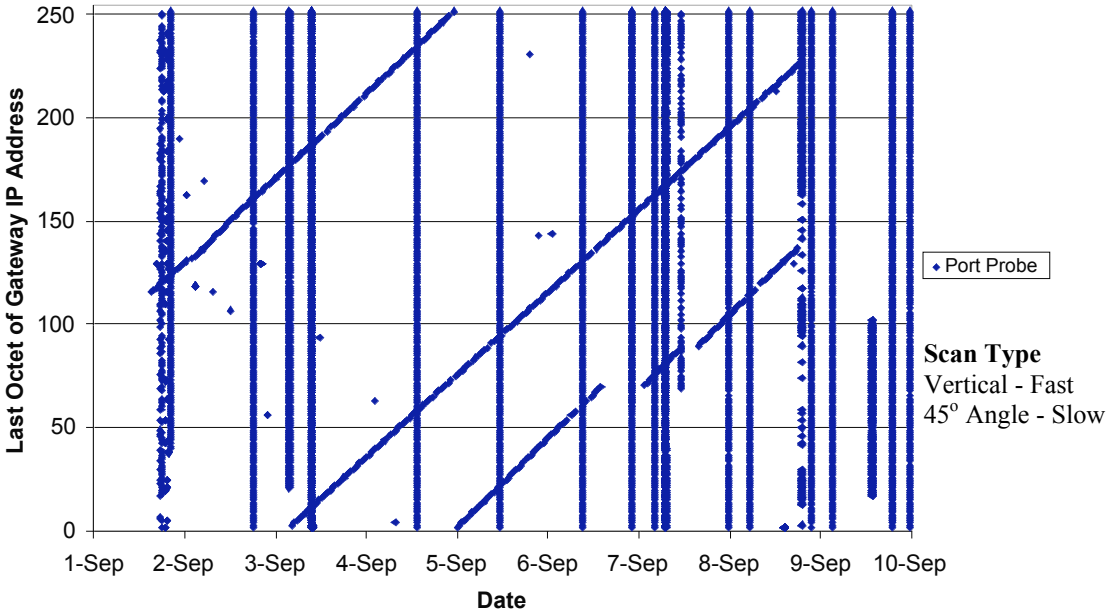

Fig 4 Ten days of gateway activity recorded by the Tracker Module

These scans are taking place over a time interval of between one minute and one hour, and will be referred to as 'fast' or 'normal' scans. Generally these are occurring as quickly as the internet connection and processor speed of the attacking machine will allow. The diagonal lines on a 45-degree angle spanning a number of days are referred to as a 'slow' scan. These will have a far larger time interval between each individual probe of anywhere up to an hour, making them more difficult to detect. Using the Tracking module it is possible to track a single IP and watch what it has done over a period of time, Fig. 5 displays the methodicalness of a port scan across the entire range of the class C address as well as the fact the scan also covered multiple ports on each of the gateways. It is this style of attack that our implementation is attempting to detect, and act upon to protect the given network.

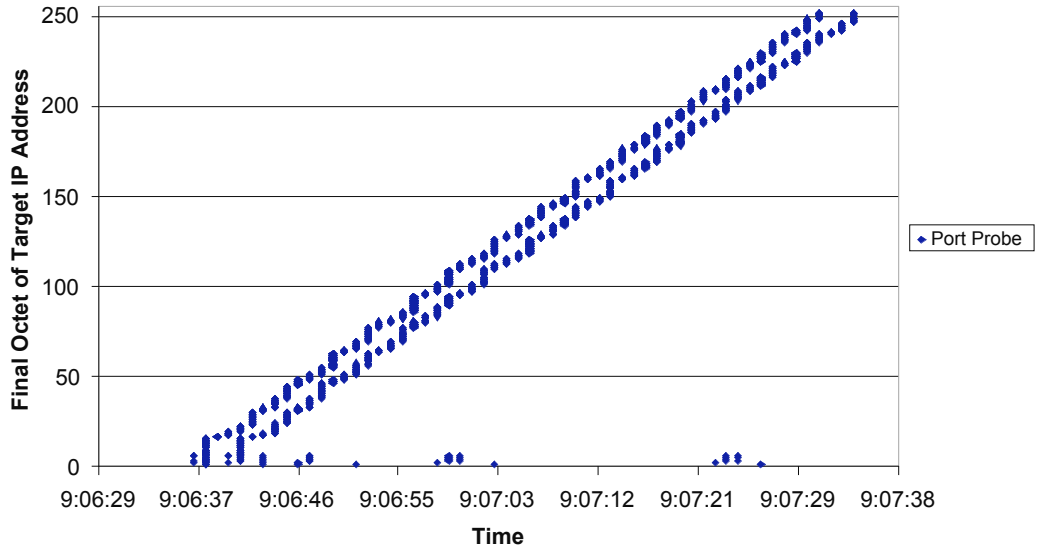

Fig. 5 .Port scan across multiple gateways and ports from a lone source IP address.

CONTINUING WORK

The implementation has progressed on to automating the process to allow it to occur in real-time, while also building in interaction between the Analysis and Tracking module. A third module has also been created called the Action module which examines the results produced by the other two and formulates a firewall rule to be sent to the gateways to provide protection from a multiple gateway attack.

The goal of the Action module is be a pre-emptive defence mechanism to provide protection the gateways on the network that have not yet been attacked by a given IP address. For this to be truly worthwhile the process needs to be efficient and operate at real-time during attacks.

The current work on the system is also aiming at discovering the optimum ban length for attackers to prevent both long-term slow scans and attackers who return after a period of inactivity. This optimum length will be one of the heuristics considered when creating the firewall rules in response to a detected attack.

RELATED WORK

There have been continuous developments and advancements in the examination of electronic audit logs since the 1980's. However this has only rarely branched into the realm of amalgamating logs across gateways to examine threats at a network wide level. Recent research in this area has occurred however with the MINDS project.

MINDS

The MINDS [12]Minnesota Intrusion Detection System) project has the objective of producing a system which will allow large scale analysis using data mining algorithms to detect attacks [12]. The MINDS system uses a combination of signature detection and anomaly detection to provide protection to the University of Minnesota network.

The MINDS system uses network traffic flow data collected from CISCO routers. This audit data is then filtered to remove extraneous entries before feature extraction collates the required information for analysis (source and destination IP's and ports, protocols, timestamp, flags). Also catalogued is derived contextual information such as the amount of traffic to a destination from a specific source. The extracted, reduced log is then run through the Attack Detection Module of MINDS using signature detection to discover any known attacks. The remaining log is then fed through the Anomaly Detection Modules that allocates a score to each connection in relation to normal traffic patterns. Connections that score highly are then further analysed by the network administrators to moderate whether or not the connection was an intrusion or a false positive. Connections that scored highly by the Anomaly Module, and are not found to be false positives by the administrators, are then further analysed to produce new signatures for emerging attacks. It is in this way the MINDS system is able to not only protect against the more common and well known attacks, but is also very strong on the detection of novel attacks, or attacks which are not yet supported by many other IDS [13, 14].

The MINDS project has been developed during the same period as our own system, and as such, there are some notable differences between the two implementations. The MINDS project does not employ threshold level heuristics in their detection mechanisms, and the system is not fully intended to be automated process.

CONCLUSION

In conclusion, this paper has demonstrated that it is possible to detect and indeed track malicious scans across a series of gateways

through the centralised analysis of audit logs. Detecting such attacks is impossible for individual gateways spread across a network as they lack the contextual information to recognise the true aims of a single trivial probe against a port.

Once the central processing and analysis has occurred, and detected malicious source IP addresses it is possible to defend the remaining network gateways from future attacks. Such pre-emptive defence mechanisms will
better equip network administrators to protect and maintain services to users within sizeable networks.

REFERENCES

[1] T. Micro, "The Real Cost of a Virus Outbreak," Trend Micro, Cupertino CA, White Paper 2002.

[2] C. P. Pfleeger and S. L. Pfleeger, *Security in computing*, 3rd Int ed. Upper Saddle River, N.J.: Prentice Hall PTR, 2003.

[3] C. Stoll, *The cuckoo's egg : tracking a spy through the maze of computer espionage*. London: Pan Books, 1991.

[4] J. Anderson, "Computer Security Threat Monitoring and Surveillance.," Fort Washington April 1980 1980.

[5] R. Kemmerer and G. Vigna, "Intrustion Detection: A Brief History and Overview," *IEEE Computer*, vol. Special publication on Security and Privacy, 2002.

[6] E. G. Amoroso, *Intrusion detection : an introduction to Internet surveillance, correlation, traps, trace back, and response*, 1st ed. Sparta, N.J.: Intrusion.Net Books, 1998.

[7] L. Heberlein, G. Dias, K. Levitt, B. Mukherjee, J. Wood, and D. Wolber, "A Network Security Monitor," *Proceedings of the 1990 IEEE Symposium on Research in Security and Privacy*, pp. 296-304, 1990.

[8] G. Holden, *Guide to Network Defence and Countermeasures*: Thomson, 2003.

[9] S. Hofmeyr, S. Forrest, and A. Somayaji, "Intrusion Detection using Sequences of System Calls," *Journal of Computer Security*, vol. 6, pp. 151/180, 1998.

[10] S. Kumar, "Classification and detection of computer intrusions.," in *Computer Science*: Purdue University, 1995, pp. 180.

[11] A. Brox, "Signature Based or Anomaly Based Intrusion Detection – The Practice and Pitfalls," *Schmagazine*, 2002.

[12] R. T. MINDS, "MINDS - Minnesota Intrusion Detection System," V. Kumar and J. Srivastava, Eds., 2004.

[13] L. Ertoz, E. Eilertson, A. Lazarevic, P. Tan, P. Dokas, J. Srivastava, and V. Kumar, "Detection and Summarization of Novel Network Attacks Using Data Mining," Technical Report 2003.

[14] L. Ertoz, E. Eilertson, A. Lazarevic, P. Tan, J. Srivastava, V. Kumar, and P. Dokas, "The MINDS - Minnesota Intrusion Detection System," in *Next Generation Data Mining*, 2004.